Infinite-message Interactive Function Computation in Collocated Networks[1]

Nan Ma
ECE Dept, Boston University
Boston, MA 02215
nanma@bu.edu

Prakash Ishwar
ECE Dept, Boston University
Boston, MA 02215
pi@bu.edu

***Abstract*—An interactive function computation problem in a collocated network is studied in a distributed block source coding framework. With the goal of computing a desired function at the sink, the source nodes exchange messages through a sequence of error-free broadcasts. The infinite-message minimum sum-rate is viewed as a functional of the joint source pmf and is characterized as the least element in a partially ordered family of functionals having certain convex-geometric properties. This characterization leads to a family of lower bounds for the infinite-message minimum sum-rate and a simple optimality test for any achievable infinite-message sum-rate. An iterative algorithm for evaluating the infinite-message minimum sum-rate functional is proposed and is demonstrated through an example of computing the minimum function of three Bernoulli sources.**

I. Introduction

In this paper, we study, using a distributed block source coding framework, an interactive function computation problem in a collocated network where nodes take turns to broadcast messages over multiple rounds. Consider a network consisting of m source nodes and a sink node. Each source node observes a discrete memoryless stationary source. The sources at different nodes are independent. The sink does not observe any source and needs to compute a samplewise function of all the sources. To achieve this objective, the nodes take turns to broadcast t messages in total. Nodes are collocated, meaning that every message is recovered at every node without error. After all the message broadcasts, the sink computes the samplewise function. The communication is said to be interactive if $t > m$.

For all finite t, a single-letter characterization of the set of all feasible coding rates (the rate region) and the minimum sum-rate was provided in [1] using traditional information-theoretic techniques. This, however, does not lead to a satisfactory characterization of the *infinite-message limit* of the minimum sum-rate as the number of messages t tends to infinity. The objective of this paper is to provide a "limit-free" characterization of the infinite-message minimum sum-rate, i.e., it does not involve taking a limit as $t \to \infty$, and also an iterative algorithm to evaluate it. This result is similar to that provided in [2], where a two-terminal interactive function computation problem was studied. The infinite-message minimum sum-rate is the fundamental limit of cooperative function computation, where potentially an infinite number of infinitesimal-rate messages can be used. While the asymptotics of blocklength, rate, quantizer step-size, and network size have been explored in the distributed source coding literature, asymptotics involving an infinite number of messages has not, to the best of our knowledge, been studied and is not well understood.

In this paper, we view the infinite-message minimum sum-rate as a functional of the joint source pmf. The main result is the characterization this functional as the least element in a partially ordered family of functionals having certain *convex-geometric* properties. This characterization does not involve taking a limit as the number of messages goes to infinity. The proof of this main result suggests an iterative algorithm for evaluating the infinite-message minimum sum-rate functional. We demonstrate this algorithm through an example of computing the minimum function of three sources.

Related interactive computation problems in various networks have been studied in [3]–[6] using the framework of communication complexity [7], [8], where computation is required to be error-free. A function computation problem in a collocated network is studied in [9] within a distributed block source coding framework, under the assumption that conditioned on the desired function, the observations of source nodes are independent. Multiround (interactive) function computation in a two-terminal network is studied in [2], [10], [11] within a distributed block source coding framework. The impact of transmission noise on function computation is considered in [12]–[14] but without a block coding rate.

The rest of this paper is organized as follows. In Sec. II, we setup the problem and recap previous results. In Sec. III we provide the main result, a "limit-free" characterization of the infinite-message minimum sum-rate. In Sec. IV we present an iterative algorithm for evaluating the minimum sum-rate functional and demonstrate it through an example.

II. Interactive Computation in Collocated Networks

A. Problem formulation

Consider a network consisting of m source nodes numbered $1, \ldots, m$, and an un-numbered sink (node). Each source node observes a discrete memoryless stationary source taking values in a finite alphabet. The sink has no source samples. For

[1]This material is based upon work supported by the US National Science Foundation (NSF) under award (CAREER) CCF–0546598 and CCF–0915389. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

each $j = 1, \ldots, m$, let $\mathbf{X}_j := (X_j(1), \ldots, X_j(n)) \in (\mathcal{X}_j)^n$ denote the n source samples which are available at node-j. In this paper, we assume sources are independent, i.e., for $i = 1, \ldots, n$, $(X_1(i), X_2(i), \ldots, X_m(i))$ are iid $p_{X^m} \in \mathcal{P}_{X^m}$ where $\mathcal{P}_{X^m} := \left\{\prod_{j=1}^m p_{X_j}\right\}$ is the set of all product pmfs on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_m$. We adopt this assumption for two reasons: (1) to isolate the impact of the structure of the desired function on the efficiency of computation, (2) to obtain an exact characterization of the optimal efficiency. The general problem where the sources are dependent across nodes is open. Let $f : \mathcal{X}_1 \times \ldots \times \mathcal{X}_m \to \mathcal{Z}$ be the function of interest at the sink and let $Z(i) := f(X_1(i), \ldots, X_m(i))$. The tuple $\mathbf{Z} := (Z(1), \ldots, Z(n))$, which denotes n samples of the samplewise function, needs be computed at the sink.

The communication is initiated by node-k. The nodes take turns to broadcast messages in t steps. In the i-th step, node-j, where $j = (k + i - 1 \mod m)$, [2] generates a message as a function of the source samples $\mathbf{X}_j$ and all the previous messages and broadcasts it. Nodes are collocated, meaning that every broadcasted message is recovered without error at every node. After t message broadcasts, the sink computes the samplewise function based on all the messages. If $t > m$, the communication is multi-round and will be called interactive.

Definition 1: A t-message distributed block source code for function computation initiated by node-k in a collocated network with parameters $(t, n, |\mathcal{M}_1|, \ldots, |\mathcal{M}_t|)$ is the tuple $(e_1, \ldots, e_t, g)$ consisting of t block encoding functions $e_1, \ldots, e_t$ and a block decoding functions g, of block-length n, where for every $i = 1, \ldots, t$, $j = (k + i - 1 \bmod m)$,

$$e_i : \left(\mathcal{X}_j\right)^n \times \bigotimes_{l=1}^{i-1} \mathcal{M}_l \to \mathcal{M}_i, \quad g : \bigotimes_{l=1}^{t} \mathcal{M}_l \to \mathcal{Z}.$$

The output of e_i, denoted by M_i, is called the i-th message. The output of g is denoted by $\widehat{\mathbf{Z}}$. For each i, $(1/n)\log_2 |\mathcal{M}_i|$ is called the i-th block-coding rate (in bits per sample).

Remark 1: (i) Each message M_i could be a null message ($|\mathcal{M}_i| = 1$). By incorporating null messages, the coding scheme described above subsumes all orders of messages transfers from m source nodes, and a t-round coding scheme subsumes a t'-round coding scheme if $t' < t$. (ii) Since the information available to the sink is also available to all source nodes, there is no advantage in terms of sum-rate to allow the sink to send any message. (iii) Although the problem studied in [1] is a special case with $k = 1$ and $t = mr$, where $r \in \mathbb{Z}^+$ is the number of rounds, the characterizations for the rate region and the minimum sum-rate in [1] naturally extend to the general problem described above.

Definition 2: A rate tuple $\mathbf{R} = (R_1, \ldots, R_t)$ is admissible for t-message function computation initiated by node-k if, $\forall \epsilon > 0$, $\exists\ \overline{n}(\epsilon, t)$ such that $\forall n > \overline{n}(\epsilon, t)$, there exists a t-message distributed block source code with parameters $(t, n, |\mathcal{M}_1|, \ldots, |\mathcal{M}_t|)$ satisfying

$$\forall i = 1, \ldots, t,\ \frac{1}{n}\log_2 |\mathcal{M}_i| \leq R_i + \epsilon, \quad \mathbb{P}(\widehat{\mathbf{Z}} \neq \mathbf{Z}) \leq \epsilon.$$

[2] $j = (k \bmod m)$ means that $j \in \{1, \ldots, m\}$ and m divides $(j - k)$.

The set of all admissible rate tuples, denoted by $\mathcal{R}_t^k$, is called the operational rate region for t-message function computation initiated by node-k. The minimum sum-rate $R_{sum,t}^k$ is given by $\min_{\mathbf{R} \in \mathcal{R}_t^k}\left(\sum_{i=1}^t R_i\right)$. The focus of this paper is on the minimum sum-rate rather than the rate region.

Remark 2: (i) We allow the number of messages t to be equal to 0 and abbreviate $R_{sum,0}^k$ to $R_{sum,0}$ because there is no message transfer and the initial-node is irrelevant. (ii) For $t < m$, function computation may be infeasible, i.e., $\mathcal{R}_t^k$ may be empty. If so, we define $R_{sum,t}^k := +\infty$. For special p_{X^m} and f, however, computation may be feasible even with $t < m$; in that case, $R_{sum,t}^k$ would be finite. (iii) For all $\tau \in \mathbb{Z}^+$, $R_{sum,t}^k \geq R_{sum,t+\tau}^k \geq 0$ holds, because the last τ messages could be null. Hence the limit $\lim_{t\to\infty} R_{sum,t}^k =: R_{sum,\infty}^k$ exists and is finite. (iv) For all $\tau \in \mathbb{Z}^+$, $R_{sum,t}^k \geq R_{sum,t+\tau}^{(k-\tau \bmod m)}$ holds, because the first τ messages could be null. It follows that $R_{sum,\infty}^k$ is independent of k and we abbreviate it to $R_{sum,\infty}$. For all finite t, however, we keep the superscript in $R_{sum,t}^k$ because this notation is convenient in the proof of Theorem 1.

For all finite t, a single-letter characterization of $\mathcal{R}_t^k$ and $R_{sum,t}^k$ was provided in Theorem 1 and Corollary 1 of [1]. This, however, does not directly lead to a satisfactory characterization of the infinite-message limit $R_{sum,\infty}$, which is a new dimension for asymptotic-analysis involving potentially an infinite number of infinitesimal-rate messages. The main contribution of this paper is a novel convex-geometric characterization of $R_{sum,\infty}$.

B. Characterization of $R_{sum,t}^k$ for finite t

Fact 1: (Characterization of $R_{sum,t}^k$ [1, Corollary 1])

$$R_{sum,t}^k = \min_{p_{U^t|X^m} \in \mathcal{P}_t^k(p_{X^m})} I(X^m; U^t), \tag{2.1}$$

where $\mathcal{P}_t^k(p_{X^m})$ is the set of all $p_{U^t|X^m}$ such that (i) $H(f(X^m)|U^t) = 0$, (ii) $\forall i \in \{1, \ldots, t\}, j = (k + i - 1 \bmod m), U_i - (U^{i-1}, X_j) - (X^{j-1}X_{j+1}^m)$, and (iii) the cardinalities of the alphabets of the auxiliary random variables U^t are upper-bounded by functions of $|\mathcal{X}_1|, \ldots, |\mathcal{X}_m|$ and t.

The Markov chain conditions in Fact 1 are equivalent to the following factorization of $p_{U^t|X^m}$:

$$p_{U^t|X^m} = p_{U_1|X_k} \cdot p_{U_2|X_{(k+1 \bmod m)}U_1} \cdot p_{U_3|X_{(k+2 \bmod m)}U^2} \cdots. \tag{2.2}$$

The cardinality bounds in Fact 1 which can be derived using the Carathéodory theorem are omitted here for clarity. Although the exact expressions of the cardinality bounds are unimportant for our discussion, a key property that needs to be highlighted is that the bound on the alphabet of U_t increases exponentially with respect to (w.r.t.) t. Therefore the dimension of the optimization problem in 2.1 explodes as t increases.

Using Fact 1, we could compute $R_{sum,t}^k$ for a large t to approximate $R_{sum,\infty}$. This is impractical because (i) the dimension of the optimization problem is large, (ii) the characterization of $R_{sum,t}^k$ does not inform us how close $R_{sum,t}^k$ is to $R_{sum,\infty}$. Alternatively, we could compute $R_{sum,t}^k$ for increasing values of t until $|R_{sum,t-1}^k - R_{sum,t}^k|$ falls below a threshold. However,

the dimensionality of the optimization problem grows exponentially with increasing values of t and there is no obvious way to reuse the computations done for evaluating $R^k_{sum,t-1}$ when evaluating $R^k_{sum,t}$. Finally, if we need to evaluate $R_{sum,\infty}$ for a different p_{X^m}, we need to repeat the entire process.

In Sec. III, we take a new fundamentally different approach. We first view $R_{sum,\infty}$ as a functional of p_{X^m} for a fixed f. Then we develop a convex-geometric blocklength-free characterization of the entire functional $R_{sum,\infty}(p_{X^m})$ which does not involve taking a limit as $t \to \infty$. This leads to a simple test for checking if a given achievable sum-rate functional of p_{X^m} coincides with $R_{sum,\infty}(p_{X^m})$. It also provides a whole new family of lower bounds for $R_{sum,\infty}$. In Sec. IV, we use the new characterization to develop an iterative algorithm for computing the functional $R_{sum,\infty}(p_{X^m})$ and $R^k_{sum,t}(p_{X^m})$ (for any finite t) in which, crudely speaking, the complexity of computation in each iteration does not grow with iteration number, and results from the previous iteration are reused in the following one. We demonstrate the iterative algorithm through an example.

III. Characterization of $R_{sum,\infty}(p_{X^m})$

A. The rate reduction functional $\rho^k_t(p_{X^m})$

If the goal is to *losslessly reproduce* the sources, the minimum sum-rate is equal to $H(X^m) = \sum_{k=1}^m H(X_k)$ because the sources are independent. The minimum sum-rate for function computation cannot be larger than that for lossless source reproduction. The reduction in the minimum sum-rate for function computation in comparison to source reproduction is given by

$$\rho^k_t := H(X^m) - R^k_{sum,t} = \max_{p_{U^t|X^m} \in \mathcal{P}^k_t(p_{X^m})} H(X^m|U^t). \tag{3.3}$$

A quantity which plays a key role in the characterization of $R_{sum,\infty}$ is ρ_0 – the "rate reduction" for zero messages (there are no auxiliary random variables in this case). Let

$$\mathcal{P}_f := \{p_{X^m} \in \mathcal{P}_{X^m} : H(f(X^m)) = 0\}.$$

Error-free computations can be performed without any message transfers if, and only if, $p_{X^m} \in \mathcal{P}_f$. Thus,

$$R_{sum,0} = \begin{cases} 0, & \text{if } p_{X^m} \in \mathcal{P}_f, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\rho_0 = \begin{cases} H(X^m), & \text{if } p_{X^m} \in \mathcal{P}_f, \\ -\infty, & \text{otherwise.} \end{cases} \tag{3.4}$$

Remark 3: If $f(x^m)$ is not constant, for all $p_{X^m} \in \mathcal{P}_f$, we have $\mathrm{supp}(p_{X^m}) \neq \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$. Such p_{X^m} can only lie on the boundary of $\mathcal{P}_{X^m}$.

Evaluating $R^k_{sum,t}$ is equivalent to evaluating the rate reduction ρ^k_t. It turns out, however, that $\rho_\infty := \lim_{t\to\infty} \rho^k_t = H(X^m) - R_{sum,\infty}$ is easier to characterize than $R_{sum,\infty}$ (see Remark **??**). The rate reduction functional is the key to the characterization.

B. Main result

Generally speaking, ρ^k_t, ρ_0, and ρ_∞ depend on p_{X^m} and f. We will fix f and view $\rho^k_t(p_{X^m})$, $\rho_0(p_{X^m})$, and $\rho_\infty(p_{X^m})$ as functionals of p_{X^m} to emphasize the dependence of p_{X^m}. Instead of evaluating $\rho_\infty(p_{X^m})$ for one particular p_{X^m} as it is done in the numerical evaluation of single-terminal and Wyner-Ziv rate-distortion functions, our approach is to characterize and evaluate the functional $\rho_\infty(p_{X^m})$ for the entire set of product distributions $\mathcal{P}_{X^m}$ rather than for one particular p_{X^m}. To describe the characterization of the functional $\rho_\infty(p_{X^m})$, it is convenient to define the following family of functionals.

Definition 3: (Marginal-distributions-concave, ρ_0-majorizing family of functionals $\mathcal{F}$) The set of marginal-distributions-concave, ρ_0-majorizing family of functionals $\mathcal{F}$ is the set of all the functionals $\rho : \mathcal{P}_{X^m} \to \mathbb{R}$ satisfying the following conditions:

1) ρ_0-majorization: $\forall p_{X^m} \in \mathcal{P}_{X^m}$, $\rho(p_{X^m}) \geq \rho_0(p_{X^m})$.
2) Concavity w.r.t. marginal distributions: For all $k \in \{1, \ldots, m\}$, with p_{X_j} held fixed for all $j \neq k$, $\rho\left(\prod_{j=1}^m p_{X_j}\right)$ is a concave function of p_{X_k}.

Remark 4: Since $\rho_0(p_{X^m}) = -\infty$ for all $p_{X^m} \notin \mathcal{P}_f$, condition 1) of Definition 3 is trivially satisfied for all $p_{X^m} \in \mathcal{P}_{X^m} \setminus P_f$ (we use the convention that $\forall a \in \mathbb{R}$, $a > -\infty$). Thus the statement that ρ majorizes ρ_0 on the set $\mathcal{P}_{X^m}$ is equivalent to the statement that ρ majorizes $H(X^m)$ on the set $\mathcal{P}_f$.

Remark 5: Condition 2) does not imply that $\rho(p_{X^m})$ is concave w.r.t. the joint pmf p_{X^m}. In fact, $\mathcal{P}_{X^m}$ is not convex.

We now state and prove the main result of this paper.

Theorem 1: (i) $\rho_\infty \in \mathcal{F}$. (ii) For all $\rho \in \mathcal{F}$, and all $p_{X^m} \in \mathcal{P}_{X^m}$, we have $\rho_\infty(p_{X^m}) \leq \rho(p_{X^m})$.

The set $\mathcal{F}$ is partially ordered w.r.t. majorization. Theorem 1 says that ρ_∞ is the least element of $\mathcal{F}$. Note that there is no parameter t which needs to be sent to infinity in this characterization of ρ_∞.

To prove Theorem 1 we will establish a connection between the t-message interactive coding problem and a $(t-1)$-message subproblem. Intuitively, to construct a t-message interactive code with initial-node k and $p_{X^m} = \prod_{i=1}^m p_{X_i}$, we need to begin by choosing the first message which corresponds to choosing the auxiliary random variable U_1. Then for each realization $U_1 = u_1$, constructing the remaining part of the code becomes a $(t-1)$-message subproblem with initial-node $k^+ := (k+1 \mod m)$ with the same desired function, but with a different joint source pmf $p_{X^m|U_1} = \prod_{i=1}^m p'_{X_i}$, where for all $i \neq k$, $p'_{X_i} = p_{X_i}$ and $p'_{X_k} = p_{X_k|U_1}$. We can repeat this procedure recursively to construct a $(t-1)$-message interactive code. After t steps of recursion, we will be left with the trivial 0-message problem.

Proof: (i) We need to verify that ρ_∞ satisfies the two conditions in Definition 3:

1) Since $\forall p_{X^m} \in \mathcal{P}_{X^m}$, $R_{sum,\infty}(p_{X^m}) \leq R_{sum,0}(p_{X^m})$, we have $\rho_\infty(p_{X^m}) \geq \rho_0(p_{X^m})$.

2) For any $k \in \{1, \ldots, m\}$, consider two arbitrary distributions $p_{X_k,0}$ and $p_{X_k,1}$, and arbitrary distributions p_{X_j} for all $j \neq k$. For $u_1 = 0, 1$, let $p_{X^m,u_1} := p_{X_k,u_1} \cdot \prod_{j=1, j\neq k}^m p_{X_j}$. For

$\lambda \in (0,1)$, let $p_{X^m,\lambda} := \lambda p_{X^m,1} + (1-\lambda) p_{X^m,0}$. We will show that $\rho_\infty(p_{X^m,\lambda}) \geq \lambda\, \rho_\infty(p_{X^m,1}) + (1-\lambda)\, \rho_\infty(p_{X^m,0})$. Let $U_1^* \sim \mathrm{Ber}(\lambda)$ and $(X^m, U_1^*) \sim p_{X^m,u_1} p_{U_1^*}(u_1)$, which imply $p_{X^m} = p_{X^m,\lambda} \in \mathcal{P}_{X^m}$ and $p_{X^m|U_1^*}(\cdot|u_1) = p_{X^m,u_1} \in \mathcal{P}_{X^m}$. For all $t \in \mathbb{Z}^+$ we have,

$$
\begin{aligned}
\rho_t^k(p_{X^m,\lambda}) &= \max_{p_{U^t|X^m} \in \mathcal{P}_t^k(p_{X^m,\lambda})} H(X^m|U^t) \\
&= \max_{p_{U_1|X_k}} \left\{ \max_{\substack{p_{U_2^t|X^m U_1}: \\ p_{U_1|X_k} p_{U_2^t|X^m U_1} \in \mathcal{P}_t^k(p_{X^m,\lambda})}} H(X^m|U^t) \right\} \\
&\overset{(a)}{\geq} \max_{\substack{p_{U_2^t|X^m U_1^*}: \\ p_{U_1^*|X_k} p_{U_2^t|X^m U_1^*} \in \mathcal{P}_t^k(p_{X^m,\lambda})}} H(X^m|U_2^t, U_1^*) \\
&\overset{(b)}{=} \lambda \cdot \max_{\substack{p_{U_2^t|X^m U_1^*}(\cdot|\cdot,1): \\ p_{U_1^*|X_k} p_{U_2^t|X^m U_1^*} \in \mathcal{P}_t^k(p_{X^m,1})}} H(X^m|U_2^t, U_1^*=1) \\
&\quad + (1-\lambda) \cdot \max_{\substack{p_{U_2^t|X^m U_1^*}(\cdot|\cdot,0): \\ p_{U_1^*|X_k} p_{U_2^t|X^m U_1^*} \in \mathcal{P}_t^k(p_{X^m,0})}} H(X^m|U_2^t, U_1^*=0) \\
&\overset{(c)}{=} \lambda\, \rho_{t-1}^{k^+}(p_{X^m,1}) + (1-\lambda)\, \rho_{t-1}^{k^+}(p_{X^m,0}).
\end{aligned} \tag{3.5}
$$

In step (a) we replaced $p_{U_1|X_k}$ with the particular $p_{U_1^*|X_k}$ defined above. Step (b) follows from the "law of total conditional entropy" with the additional observations that conditioned on $U_1^* = u_1$, $p_{X^m|U_1^*}(\cdot|u_1) = p_{X^m,u_1}$ and $H(X^m|U_2^t, U_1^* = u_1)$ only depends on $p_{U_2^t|X^m U_1^*}(\cdot|\cdot, u_1)$. Step (c) is due to the observation that for a fixed $p_{U_1^*|X_k}$, conditioned on $U_1^* = u_1$, $p_{U_1^*|X_k} p_{U_2^t|X^m U_1^*} \in \mathcal{P}_t^k(p_{X^m,u_1})$ iff $p_{U_2^t|X^m U_1^*} \in \mathcal{P}_{t-1}^{k^+}(p_{X^m,u_1})$. Now send t to infinity in both the left and right sides of (3.5). Since $\lim_{t\to\infty} \rho_t^k = \lim_{t\to\infty} \rho_t^{k^+} = \rho_\infty$, we have $\rho_\infty(p_{X^m,\lambda}) \geq \lambda\rho_\infty(p_{X^m,1}) + (1-\lambda)\rho_\infty(p_{X^m,0})$. Therefore, ρ_∞ satisfies condition 2) in Definition 3. Thus, $\rho_\infty \in \mathcal{F}$.

(ii) It is sufficient to show that: $\forall \rho \in \mathcal{F}$, $\forall p_{X^m} \in \mathcal{P}_{X^m}$, $\forall t \in \mathbb{Z}^+ \bigcup \{0\}$, $\forall k \in \{1,\ldots,m\}$, $\rho_t^k(p_{X^m}) \leq \rho(p_{X^m})$. We prove this by induction on t. For $t = 0$, the result is true by condition 1) in Definition 3. Assume that for an arbitrary $t \in \mathbb{Z}^+$, $\rho_{t-1}^k(p_{X^m}) \leq \rho(p_{X^m})$ holds. We will show that $\rho_t^k(p_{X^m}) \leq \rho(p_{X^m})$ holds.

$$
\begin{aligned}
\rho_t^k(p_{X^m}) &= \max_{p_{U^t|X^m} \in \mathcal{P}_t^k(p_{X^m})} H(X^m|U^t) \\
&= \max_{p_{U_1|X_k}} \left\{ \max_{\substack{p_{U_2^t|X^m U_1}: \\ p_{U_1|X_k} p_{U_2^t|X^m U_1} \in \mathcal{P}_t^k(p_{X^m})}} H(X^m|U^t) \right\} \\
&\overset{(d)}{=} \max_{p_{U_1|X_k}} \left\{ \sum_{u_1 \in \mathrm{supp}(p_{U_1})} p_{U_1}(u_1) \left\{ \max_{\substack{p_{U_2^t|X^m U_1}(\cdot|\cdot,u_1): \\ p_{U_1|X_k} p_{U_2^t|X^m U_1} \in \mathcal{P}_t^k(p_{X^m|U_1}(\cdot|u_1))}} H(X^m|U_2^t, U_1 = u_1) \right\} \right\} \\
&\overset{(e)}{=} \max_{p_{U_1|X_k}} \left\{ \sum_{u_1 \in \mathrm{supp}(p_{U_1})} p_{U_1}(u_1)\, \rho_{t-1}^{k^+}(p_{X^m|U_1}(\cdot|u_1)) \right\} \\
&\overset{(f)}{\leq} \max_{p_{U_1|X_k}} \left\{ \sum_{u_1 \in \mathrm{supp}(p_{U_1})} p_{U_1}(u_1)\, \rho(p_{X^m|U_1}(\cdot|u_1)) \right\} \\
&\overset{(g)}{=} \max_{p_{U_1|X_k}} \left\{ \sum_{u_1 \in \mathrm{supp}(p_{U_1})} p_{U_1}(u_1)\, \rho(p_{X_k|U_1}(\cdot|u_1)\; p_{X^{k-1}X_{k+1}^m}) \right\} \\
&\overset{(h)}{\leq} \max_{p_{U_1|X_k}} \left\{ \rho\left(\sum_{u_1 \in \mathrm{supp}(p_{U_1})} p_{U_1}(u_1) p_{X_k|U_1}(\cdot|u_1)\; p_{X^{k-1}X_{k+1}^m} \right) \right\} \\
&= \rho(p_{X^m}).
\end{aligned} \tag{3.6}
$$

The reasoning for steps (d) and (e) are similar to those for steps (b) and (c) respectively in the proof of part (i) (see equation array (3.5)). In step (e) we need to use the fact that $p_{X^m|U_1}(\cdot|u_1) \in \mathcal{P}_{X^m}$, which is due to (2.2) and the assumption that $p_{X^m} \in \mathcal{P}_{X^m}$. Step (f) is due to the induction hypothesis $\rho_{t-1}^k(p_{X^m}) \leq \rho(p_{X^m})$ for all k. Step (g) is due to the Markov chain $U_1 - X_k - (X^{k-1}X_{k+1}^m)$ and because X_k and $(X^{k-1}X_{k+1}^m)$ are independent. Step (h) is Jensen's inequality applied to $\rho(p_{X_k} \cdot p_{X^{k-1}X_{k+1}^m})$ which is concave w.r.t. p_{X_k}. ∎

Since every $\rho \in \mathcal{F}$ gives an upper bound for ρ_∞, $(H(X^m) - \rho)$ gives a lower bound for $R_{sum,\infty}$. This fact provides a method for testing if an achievable sum-rate functional is optimal. If $R^*(p_{X^m})$ is an achievable sum-rate functional then $\forall p_{X^m} \in \mathcal{P}_{X^m}$, $R^*(p_{X^m}) \geq R_{sum,\infty}(p_{X^m})$. If it can be verified that $\rho^* := (H(X^m) - R^*) \in \mathcal{F}$, then by Theorem 1, $R^* = R_{sum,\infty}$.

IV. Iterative algorithm

Although Theorem 1 provides a characterization of ρ_∞ and $R_{sum,\infty}$ that is not obtained by taking a limit, it does not directly provide an algorithm to evaluate $R_{sum,\infty}$. To efficiently represent and search for the least element of $\mathcal{F}$ is nontrivial because each element is a functional; not a scalar. The proof of Theorem 1, however, inspires an iterative algorithm for evaluating $R_{sum,t}^k$ and $R_{sum,\infty}$.

Equation (3.6) states that $\rho_t^k(p_{X^m})$ is the maximum value of $\rho \in \mathbb{R}$ such that (p_{X^m}, ρ) is a finite convex combination of $\{(p_{X^m|U_1}(\cdot|u_1), \rho_{t-1}^{k^+}(p_{X^m|U_1}(\cdot|u_1))\}_{u_1 \in \mathrm{supp}(p_{U_1})}$, where $p_{X^m}(\cdot)$ and $p_{X^m|U_1}(\cdot|u_1)$ have the same marginal distributions p_{X_j} for all $j \neq k$ and differ only on p_{X_k}. Now we fix the marginal distributions p_{X_j} for all $j \neq k$, and consider the hypograph of $\rho_{t-1}^{k^+}$ w.r.t. p_{X_k}: $\mathrm{hyp}_{p_{X_k}} \rho_{t-1}^{k^+} := \{(p_{X_k}, \rho) : \rho \leq \rho_{t-1}^{k^+}(\prod_{i=1}^m p_{X_i})\}$. Due to (3.6), the convex hull of $\mathrm{hyp}_{p_{X_k}} \rho_{t-1}^{k^+}$ is $\mathrm{hyp}_{p_{X_k}} \rho_t^k$. This relation enables us to evaluate ρ_t^k from $\rho_{t-1}^{k^+}$: fixing p_{X_j} for all $j \neq k$, ρ_t^k is the least concave functional w.r.t. p_{X_k} that majorizes $\rho_{t-1}^{k^+}$. In the convex optimization literature, $(-\rho_t^k)$ is called the double Legendre-Fenchel transform or convex biconjugate of $(-\rho_{t-1}^{k^+})$ [15]. We have the following iterative algorithm.

Algorithm to evaluate $R_{sum,t}^k$

- **Initialization:** For all $k = 1,\ldots,m$, define $\rho_0^k(p_{X^m}) = \rho_0(p_{X^m})$ by equation (3.4) for all p_{X^m} in $\mathcal{P}_{X^m} = \left\{\prod_{i=1}^m p_{X_i}\right\}$.
- **Loop:** For $\tau = 1$ through t do the following.
  For every $k = 1,\ldots,m$ do the following.
  For every set of marginal distributions $\{p_{X_j}\}_{j=1,j\neq k}^m$ do the following.
  - Construct $\mathrm{hyp}_{p_{X_k}} \rho_{\tau-1}^{k^+}$.
  - Let ρ_τ^k be the upper boundary of the convex hull of $\mathrm{hyp}_{p_{X_k}} \rho_{\tau-1}^{k^+}$.

- **Output:** $R^k_{sum,t}(p_{X^m}) = H(X^m) - \rho^k_t(p_{X^m})$.

To make numerical computation feasible, $\mathcal{P}_{X^m}$ has to be discretized. Once discretized, however, in each iteration, the amount of computation is the same and is fixed by the discretization step-size. Also note that results from each iteration are reused in the following one. Therefore, for large t, the complexity to compute $R^k_{sum,t}$ grows linearly w.r.t. t.

$R_{sum,\infty}$ can also be evaluated to any precision, in principle, by running this iterative algorithm for $t = 1, 2, \ldots$, until some stopping criterion is met, e.g., the maximum difference between ρ^k_{t-1} and ρ^k_t on $\mathcal{P}_{X^m}$ falls below some threshold. Developing stopping criteria with precision guarantees requires some knowledge of the rate of convergence which is not established in this paper and will be explored in future work. When the objective is to evaluate $R_{sum,\infty}(p_{X^m})$ for all pmfs in $\mathcal{P}_{X^m}$, this iterative algorithm is much more efficient than using (2.1) to solve for $R^k_{sum,t}$ for each p_{X^m} for $t = 1, 2, \ldots$, an approach which follows the definition of $R_{sum,\infty}$ literally as the limit of $R^k_{sum,t}$ as $t \to \infty$. Our iterative algorithm is based on Theorem 1 which is a characterization of $R_{sum,\infty}$ *without taking a limit involving* t.

Example: (MIN function) Take $m = 3$ nodes. $X_i \sim \text{Ber}(p_i)$. $f(x^3) = \min_{i=1,2,3} x_i$. The joint pmf p_{X^3} is parameterized by $\mathbf{p} = (p_1, p_2, p_3) \in [0,1]^3$. It is easy to see that

$$\rho_0(\mathbf{p}) = \begin{cases} \sum_{i=1}^3 h_2(p_i), & \text{if } \mathbf{p} \in \mathcal{P}_f, \\ -\infty, & \text{otherwise,} \end{cases}$$

where $\mathcal{P}_f = \{\mathbf{p} : p_1 = p_2 = p_3 = 1 \text{ or } p_1p_2p_3 = 0\}$.

Now let us fix p_{X_1} and p_{X_2} and apply the convex biconjugate operation on ρ_0 w.r.t. p_{X_3} to obtain ρ^3_1. Specifically, for every fixed (p_1, p_2), we focus on ρ_0 on the line segment $\{p_1\}\times\{p_2\}\times[0,1]$ and convexify $\text{hyp}_{p_3}\rho_0 = \{(p_3, \rho) : \rho \le \rho_0(p_1, p_2, p_3)\}$ to obtain $\text{hyp}_{p_3}\rho^3_1$. Then we repeat this procedure but applying the convexification operation w.r.t. p_{X_2}, p_{X_1}, etc to obtain ρ^2_2, ρ^1_3, etc. In numerical computation, $\mathbf{p}$ takes values on a discrete grid where p_1, p_2, p_3 are multiples of a finite step size Δ. The convexification operation involves finding a convex hull of a finite number of points in a plane.

As we decrease Δ and increase t, ρ_t approximates ρ_∞. Fig. 1 shows $\rho_t(1/2, 1/2, 1/2)$ for different t and Δ. For each Δ, ρ_t converges as t increases. For a small enough Δ (fine enough discretization), the limit represents the actual value of $\rho_\infty(1/2, 1/2, 1/2)$. Notice that for a small enough Δ, ρ_t keeps increasing as t grows, which means there is always an improvement for using more messages.

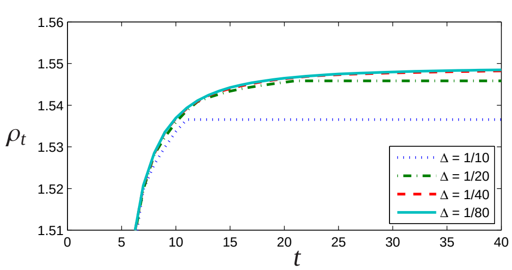

Fig. 1. $\rho_t(1/2, 1/2, 1/2)$ *for different step sizes* Δ

Fig. 2 shows the plots of the rate reduction function with $t = 40$ for four values of

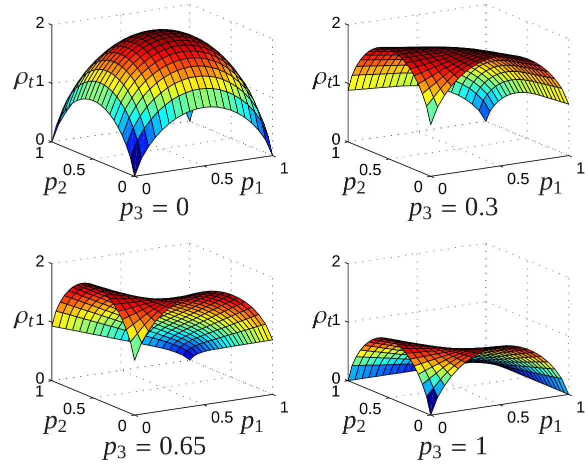

Fig. 2. *Rate reduction function for* $t = 40$

References

[1] N. Ma, P. Ishwar, and P. Gupta, "Information-theoretic bounds for multiround function computation in collocated networks," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, Jun. 28–Jul. 3, 2009, pp. 2306 – 2310.

[2] N. Ma and P. Ishwar, "Infinite-message distributed source coding for two-terminal interactive computing," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 30–Oct. 7, 2009.

[3] A. Giridhar and P. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Sel. Areas of Commun.*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[4] S. Subramanian, P. Gupta, and S. Shakkottai, "Scaling bounds for function computation over large networks," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 24–29, 2007, pp. 136–140.

[5] R. Appuswami, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network coding for computing," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 23–26, 2008, pp. 1–6.

[6] H. Kowshik and P. R. Kumar, "Zero-error function computation in sensor networks," in *Proc. IEEE Conf. Decision Control*, Shanghai, China, Dec. 16–18, 2009.

[7] A. C. Yao, "Some complexity questions related to distributed computing," in *Proc. 11th Annu. ACM Symp. Theory of Computing*, Atlanta, GA, Apr. 30–May 2, 1979, pp. 209–213.

[8] E. Kushilevitz and N. Nisan, *Communication Complexity*. Cambridge: Cambridge University Press, 1997.

[9] V. Prabhakaran, K. Ramchandran, and D. Tse, "On the role of interaction between sensors in the CEO problem," in *Proc. 42th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sep. 29–Oct. 1, 2004.

[10] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[11] N. Ma and P. Ishwar, "Two-terminal distributed source coding with alternating messages for function computation," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, Jul. 6–11, 2008, pp. 51–55.

[12] R. Gallager, "Finding parity in a simple broadcast network," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 176–180, Mar. 1988.

[13] L. Ying, R. Srikant, and G. Dullerud, "Distributed symmetric function computation in noisy wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4826–4833, Dec. 2007.

[14] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[15] R. T. Rockafellar, *Convex Analysis*. Princeton: Princeton University Press, 1970.